Stock Returns and Real Growth: A Bayesian Nonparametric Approach

Qiao Yang*

June 2019

Abstract

This study constructs a Bayesian nonparametric model to investigate whether stock market returns predict real economic growth. Unlike earlier studies, our use of an infinite hidden Markov model enables parameters to be time-varying across an infinite number of Markov-switching states estimated from data rather than fixed like a prior. Our model exhibits significantly greater accuracy in out-of-sample density forecasts. We uncover strong evidence of the time-varying power of lagged stock returns to predict economic growth.

Key words: hierarchical Dirichlet process prior, beam sampling, Markov-switching, MCMC,

JEL: C58, C14, C22, C11

*ShanghaiTech University, yangqiao@shanghaitech.edu.cn. I thank John Maheu, Martin Burda, and Yuanyuan Wan for their valuable comments. I am grateful for comments from the associate editor and referee. I thank the participants attending Midwest econometrics annual meeting, SBIES, and seminar participants at University of Toronto. This project is financially supported by ShanghaiTech start-upf fund.

1 Introduction

Establishing a connection between stock market returns and real economic growth is an important empirical undertaking[1]. This study examines that connection through a flexible Bayesian nonparametric approach. First, it applies an infinite hidden Markov model (IHMM) to study relations between stock returns and real economic growth. The new model facilitates research by combining linear and nonlinear relationships and surpasses existing benchmarks capturing data dynamics. Second, results document robustly that lagged stock market returns have significant time-varying power to predict real economic growth. Third, we intensively investigate lag structures and model selection through density forecasts, which has received little scholarly attention.

Fama (1990) and Schwert (1990) use linear regression find that lagged stock market returns correlate significantly with future real economic growth. Extending their work via cointegration tests and error-correction model, Choi et al. (1999) find only weak evidence that lagged stock market returns predict economic growth. Lee (1992) and Hassapis & Kalyvitis (2002) find significant correlations between lagged stock market returns and future economic growth by applying vector autoregression (VAR). Kanas & Ioannidis (2010) apply a nonlinear Markov-switching model with two states and find a nonlinear correlation between lagged stock market returns and future economic growth. Kim & In (2003) find via spectral and wavelet analysis that the power of lagged stock market returns to predict economic growth is likely time-varying. Hamilton & Lin (1996) suggest that hidden state variables primarily govern the dynamics of stock market returns and real economic growth.

Previous literature indicates a flexible model that allows time-varying parameters is necessary to capture changes in their relationships. Simple Markov-switching or VAR are insufficient. Studies also examine the predictive power of lagged stock returns through point forecasts or in-sample correlation[2]. The literature devotes scant attention to effects of different model and lag structures.

This paper introduces an IHMM as a mechanism to determine whether and how significantly lagged stock market returns predict economic growth. In contrast to Markov-switching with fixed states, IHMM allows an unbounded transition matrix to infer the number of states from data rather than fixing them as priors. Applying IHMM to excess monthly returns on the S&P500 and growth rates of industrial production produces substantially more accurate out-of-sample density forecasts than conventional benchmarks. Our model also captures the market-economy relationship through structural breaks as well as recurring states. After extensive studies of lag structures and model selection, our results convincingly suggest that lagged stock returns significantly predict

[1] Fama (1990), Schwert (1990), Choi et al. (1999) and Kanas & Ioannidis (2010) suggest real stock returns lead changes in real economic activity using dividend discount valuation or consumption capital asset pricing model.

[2] Point forecasts e.g., root mean squared forecast error (RMSFE) examines the center of a predictive distribution and ignore other quantiles (e.g., tails of predictive distributions). A density forecast evaluates the entire predictive distribution, which is critical when heteroscedasticity is present. In-sample tests (t-test) do not meaningfully examine predictive power.

economic growth in a time-varying manner[3].

This paper proceeds as follows. Section 2 introduces benchmark models. Section 3 discusses the IHMM. Section 4 describes how to compute various density forecast measurements. Section 5 explores empirical results, Section 7 concludes.

2 Benchmark Models

We investigate several benchmark model to test how well they demonstrate that stock returns predict economic growth while allowing heteroscedasticity and time-varying parameters. One category of benchmarks adopts univariate model constructed via autoregression (AR), generalized autoregressive conditional heteroscedasticity (GARCH), stochastic volatility (SV), and Markov-switching (MS-AR). Another category employs multivariate models. We investigate VAR, MS-VAR, and BEKK models as benchmarks.

Besides examining models, we study lags up to four months for stock returns and real growth to clarify the predictive power of stock returns. Hamilton & Lin (1996) suggest only one lag is necessary, but their approach relies solely on point forecasts. We investigate predictability through density forecasts in which r_t and g_t respectively represent stock market return and economic growth rates at time t. Details of each benchmark model are outlined in the following.

AR implicitly assumes one state governs an entire time series. Fama (1990) and Schwert (1990) use a similar specification. We revisit this specification under the Bayesian approach.

$$g_t = c + \sum_{j=1}^{q} a_j g_{t-j} + \sum_{j=1}^{p} b_j r_{t-j} + \sigma e_t \quad e_t \overset{iid}{\sim} N(0,1), \tag{1}$$

GARCH-AR models facilitate studying the effects of stock returns when heteroscedasticity is introduced, making it possible to tell if greater accuracy of density forecasts arises from heteroscedasticity rather than stock returns. Let σ in AR become time-varying σ_t in GARCH-AR as indicated in Eq.(2):

$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2}$$

SV-AR is an approach to modeling heteroscedasticity, that allows extra randomness in conditional variance with respect to GARCH. The σ_t thereafter becomes $\exp(\frac{h_t}{2})$, and h_t changes in the following way:

$$h_t = \omega + \alpha h_{t-1} + \sigma_\nu \epsilon_t \quad \epsilon_t \overset{iid}{\sim} N(0,1) \tag{3}$$

A two-state MS-AR model allows the predictive power of explanatory variables and conditional variance to be regime-dependent[4].

[3]Existing studies claim such predictive relationships but provide no solid evidence.

[4]Hamilton & Lin (1996) and Kanas & Ioannidis (2010) use the same model.

$$g_t = c_{s_t} + \sum_{j=1}^{q} a_{js_t} g_{t-j} + \sum_{j=1}^{p} b_{js_t} r_{t-j} + \sigma_{s_t} e_t \quad e_t \overset{iid}{\sim} N(0,1) \tag{4a}$$

$$s_t | s_{t-1} \sim \Pi_{s_{t-1}} \quad s_t \in \{1,2\} \tag{4b}$$

in which Π is the first order Markov transition matrix with two dimensions, and $\Pi_{s_{t-1}}$ represents the s_{t-1} row of transition matrix of Π[5].

Lee (1992) and Hassapis & Kalyvitis (2002) ues VAR to model stock market returns and real economic growth jointly[6]. We apply the independent Normal Wishart prior[7].

$$\mathbf{y_t} = \mathbf{c} + \sum_{j=1}^{q} \mathbf{a_j} g_{t-j} + \sum_{j=1}^{p} \mathbf{b_j} r_{t-j} + \mathbf{e_t} \quad \mathbf{e_t} \overset{iid}{\sim} MN(0, \Sigma) \tag{5}$$

BEKK-VAR is a multivariate version of GARCH-AR[8]. The Σ in VAR becomes time-varying Σ_t and changes in the following way:

$$\Sigma_t = C'C + A'\mathbf{e_{t-1}}\mathbf{e'_{t-1}}A + B'\Sigma_{t-1}B \tag{6}$$

Applying restrictions assures Σ_t is always positive definite[9].

The two-state MS-VAR which allows a unique Markov state to govern both time series[10].

$$\mathbf{y_t} = \mathbf{c_{s_t}} + \sum_{j=1}^{q} \mathbf{a_{js_t}} g_{t-j} + \sum_{j=1}^{p} \mathbf{b_{js_t}} r_{t-j} + \mathbf{e_t} \quad \mathbf{e_t} \overset{iid}{\sim} MN(0, \Sigma_{s_t}) \tag{7a}$$

$$s_t | s_{t-1} \sim \Pi_{s_{t-1}} \quad s_t \in \{1,2\} \tag{7b}$$

We tested univariate benchmark models to discover how lagged stock returns correlate with future economic growth through the right side of the equation. The multivariate approach facilitates the studying of effects of lagged and simultaneous co-movement in stock returns.

[5]The autoregression has the same priors as the AR model. The prior for the transition matrix is a Dirichlet distribution. Albert & Chib (1993) reference this sampling approach.

[6]Let the $\mathbf{y_t} = \{g_t, r_t\}'$, $\mathbf{e_t} = \{e_{gt}, e_{rt}\}'$. Parameters sets are denoted as $\mathbf{c} = \{c_g, c_r\}'$, $\mathbf{a_j} = \{a_{gj}, a_{rj}\}'$, $\mathbf{b_j} = \{b_{gj}, b_{rj}\}'$ and Σ be the variance-covariance matrix.

[7]Koop & Korobilis (2009)

[8]We sample the BEKK with independent MH with random walk.

[9]C' is a low triangle matrix with a positive diagonal. The diagonal elements of A' and B' are positive.

[10]Let the parameter set be denoted $\mathbf{c_{s_t}} = \{c_{gs_t}, c_{rs_t}\}'$, $\mathbf{a_j} = \{a_{gjs_t}, a_{rjs_t}\}'$, $\mathbf{b_j} = \{b_{gjs_t}, b_{rjs_t}\}'$. The prior is independent Normal Wishart. The prior of Π is a Dirichlet distribution.

3 Infinite Hidden Markov Model (IHMM)

IHMM is based on a Bayesian nonparametric prior introduced by Beal et al. (2002). Compared to finite Markov-switching models, such as MS-AR and MS-VAR, IHMM extends the transition probability matrix from a finite to an infinite dimension. That allows the model to learn regime dynamics through data rather than affixing it as a prior. IHMM builds on priors of the hierarchical Dirichlet process (HDP), an extension of the Dirichlet process (DP)[11].

3.1 Infinite Hidden Markov Model (IHMM)

Our use of IHMM parallels its application in other empirical fields[12]. But we focus on the predictive relationship of stock returns on real growth. State variable s_t follows a first order Markov with an infinite transition matrix, such as $s_t \in \{1, 2, 3, \dots\}$. An element in Π_j, π_{ji} represents the probability of moving from state j to state i. Γ_j is a jth row of the transition matrix of Π.

$$\Gamma = \{\gamma_j\}_{j=1}^{\infty} | \eta \sim Stick(\eta) \quad j = 1, 2, \dots, \tag{8a}$$

$$\Pi_j | \alpha, \Gamma \overset{iid}{\sim} Stick2(\alpha, \Gamma), \quad \theta_i \overset{iid}{\sim} H \tag{8b}$$

$$s_t | s_{t-1}, \Pi_{s_{t-1}} \sim \Pi_{s_{t-1}}, \quad y_t | s_t, \Theta \sim F(y_t | \theta_{s_t}). \quad t = 1, \dots, T \tag{8c}$$

In Eq.(8), two DPs construct the prior for a Markov transition probability matrix with infinite dimensions[13]. The θ_i denotes the parameter set corresponding to the ith state. $Stick(\eta)$ is like an oracle that governs the number of states introduced to the data. H represents priors over parameter space Θ. $F(.)$ is the conditional density function for observation y_t. Combinations of η and α can enforce different prior beliefs

[11] Recent econometric studies employ DP in numerous empirical applications. See Jensen & Maheu (2010), Jensen & Maheu (2013), Song (2013) and others. The Appendix defines HDP and DP in details.

[12] Song (2013) applies IHMM to U.S. real interest rates. Jochmann (2015)to U.S. inflation rates, Dufays (2015) to stock volatilities, and Maheu & Yang (2016) to U.S. three-month T-bill rates. Shi & Song (2015) employ it to detect speculative bubbles in the NASDAQ. Hou (2017) introduces a multivariate version of IHMM to study macroeconomic forecasts.

[13] Teh et al. (2006) introduced HDP. Maheu & Yang (2016) first defined $Stick2(\alpha, \Gamma)$ as follow:

$$\pi_{ji} = \hat{\pi}_{ji} \prod_{l=1}^{i-1} (1 - \hat{\pi}_{jl}), \qquad \hat{\pi}_{ji} \overset{iid}{\sim} Beta\left(\alpha\gamma_i, \alpha\left(1 - \sum_{l=1}^{i} \gamma_l\right)\right),$$

Sethuraman (1994) defines $Stick(\eta)$ in this way:

$$\gamma_j = v_j \prod_{l=1}^{j-1} (1 - v_l), \quad v_j \overset{iid}{\sim} Beta(1, \eta)$$

on the dimensions of Markov transitions. That is, larger values of α and η allow higher possibilities of visiting a new state. To reduce effect of priors on forecasting performance and to allow the model to better explore state spaces, we place a hyper prior on α and η. Doing so lets us infer them from the data rather than keeping them fixed.

$$\eta \sim Gamma(\chi_1, \chi_2) \quad \alpha \sim Gamma(\chi_3, \chi_4) \tag{9}$$

Unlike finite Markov-switching models, the unbounded transition matrix allows the conditional distribution of y_t to be constructed by any number of Gaussian mixture components. This feature significantly weakens the influence of distribution assumptions on the innovation. For instance, with any given form of the error term, IHMM can always approximate it by adjusting the number of mixture components and corresponding parameters[14].

3.2 IHMM-AR

The univariate version of the IHMM is denoted as IHMM-AR.

$$g_t = c_{s_t} + \sum_{j=1}^{q} a_{js_t} g_{t-j} + \sum_{j=1}^{p} b_{js_t} r_{t-j} + \sigma_{s_t} e_t \quad e_t \overset{iid}{\sim} N(0,1) \tag{11a}$$

$$s_t | s_{t-1} \sim \Pi_{s_{t-1}} \quad s_t \in \{1, , \ldots, \infty\} \tag{11b}$$

$$\Gamma | \eta \sim Stick(\eta), \quad \Pi_j | \alpha, \Gamma \overset{iid}{\sim} Stick2(\alpha, \Gamma), \quad j = 1, 2, \ldots, \tag{11c}$$

Let $\vartheta = (c, a_1, \ldots, a_q, b_1, \ldots, b_p)$ and the priors be $\vartheta \sim MN(d_0, D_0)$ and $\frac{1}{\sigma^2} \sim Gamma(\nu_0, \lambda_0)$. Priors for the autoregressive element in IHMM-AR are the same as for AR. Given the characteristics of Markov-switching models, we could cluster parameters within each state and update the priors. Song (2013) introduces this hierarchical priors approach. Maheu & Yang (2016) show significant improvements in accuracy of density forecasts by applying it[15].

[14]Following is the conditional distribution of y_{t+1},

$$f(y_{t+1} | \Theta, s_t) = \sum_{k=1}^{\infty} \pi_{s_t k} p(y_{t+1} | \theta_k) \tag{10}$$

[15]The following hierarchical priors also can be applied to finite Markov-switching model.

$$d_0 \sim MN(d_1, D_1) \qquad D_0^{-1} \sim Wishart(\tau_1, H_1)$$
$$\nu_0 \sim Gamma(\nu_1, \nu_2) \qquad \lambda_0 \sim Gamma(\lambda_1, \lambda_2),$$

where $Wishart(\tau_1, H_1)$ denotes the Wishart distribution with $\tau_1 \geq dim(\vartheta)$ as degree of freedom and H_1 is the scale matrix with the same dimension as H_1^{-1}.

3.3 IHMM-VAR

The IHMM-VAR for modeling stock market returns and real economic growth jointly is written similar to IHMM-AR but replaces Eq.(11a) with the following[16]:

$$\mathbf{y_t} = \mathbf{c_{s_t}} + \sum_{j=1}^{q} \mathbf{a_{js_t}} g_{t-j} + \sum_{j=1}^{p} \mathbf{b_{js_t}} r_{t-j} + \mathbf{e_t} \quad \mathbf{e_t} \overset{iid}{\sim} MN(0, \Sigma_{s_t})$$

The key difference between IHMM-AR and IHMM-VAR is that former considers only lagged relationships and the latter considers the simultaneous co-movement on top of the lagged relationships.

4 Predictive Likelihood

Predictive likelihood measures the accuracy of out-of-sample density forecasts. It evaluates the entire predictive distribution. Unlike point forecasts, which capture only its center. Thus, it is no surprise these method deliver contradictory outcomes[17]. Eq.(15) represents predictive likelihood.

$$\rho(y_{T+1}|y_{1:T}) = \int_{\Theta} f(y_{T+1}|\theta, y_{1:T})\rho(\theta|y_{1:T})d\theta, \qquad \theta \in \Theta \tag{15}$$

where marginalization is taken with respect to $\rho(\theta|y_{1:T})$; it exposes the predictive posterior distribution of θ, and the parameters of interest. Observations are denoted as $y_{1:T}$. Eq. (15) also can be used to evaluate model fit as the computation of predictive likelihood incorporates parameter uncertainties. For computing log-predictive likelihood (LPL), the first 10,000 MCMC draws are burn-in, and the next 20,000 are for predictive inference[18]. Section 4 introduces the computation of the predictive likelihood of IHMM-AR; the same principle is applied to the IHMM-VAR with slight variations in notation.

Let $\{\vartheta^{(l)_i}, \sigma_i^{(l)}, \Pi^{(l)}, s_{1:T}^{(l)}, K^{(l)}, \xi^{(l)}\}$[19] be the lth posterior draw for $i \in \{1, \ldots, K^{(l)}\}$, and M is the total number of MCMC draws used to forecast inference. Because we are

[16]Let $\vartheta = (\mathbf{c}, \mathbf{a_1}, \ldots, \mathbf{a_q}, \mathbf{b_1}, \ldots, \mathbf{b_p})$ and,

$$\vartheta \sim MN(d_0, D_0) \quad \Sigma^{-1} \sim Wishart(\tau_0, H_0) \tag{13}$$

Hierarchical priors for Eq. (13) are,

$$d_0 \sim MN(d_1, D_1) \quad D_0^{-1} \sim Wishart(\tau_1, H_1) \tag{14a}$$

$$\tau_0 \sim Gamma(\zeta_1, \zeta_2)I(\tau_0 \geq 2) \quad H_0 \sim InvWhishart(\tau_2, H_2) \tag{14b}$$

These hierarchical priors can be applied to the MS2-VAR of Eq.(7).

[17]Maheu & Yang (2016)

[18]We use a recursive method of predictive inference, The last draw for predicting $T+1$ is the first for predicting $T+2$.

[19]See Section 3.2 for notation.

interested only in the potential gain for predicting real economic growth, the predictive likelihood of g_{T+1} is the following:

1. For each $l$the MCMC draw, simulate a state variable for $s_{T+1}^{(l)}$ given $s_T^{(l)}$ according to $\Pi_{s_T}^{(l)}$.

2. If $s_{T+1}^{(l)} \leq K^{(l)}$, which suggests g_{T+1} belongs to existing states, set $(\vartheta_{s_{T+1}}^{(l)}, \sigma_{s_{T+1}}^{(l)}) \equiv (\vartheta_k^{(l)}, \sigma_k^{(l)})$, where $k \in \{1, \ldots, K^{(l)}\}$. Otherwise, it is implied that g_{T+1} belongs to a new state, such as $(\vartheta_k^{(l)}, \sigma_k^{(l)}) \sim H(\xi^{(l)})$, where $H(\xi^{(l)})$ represents Eq.(13).

$$p(g_{T+1}|g_{1:T}, r_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} N(g_{T+1}|\vartheta_k^{(l)}, \sigma_k^{2(l)})$$

The LPL is a measurement for model selection, and it evaluates accuracy of forecasts based on a selected out-of-sample period. Let l_1 and l_2 indicate the beginning and end of the out-of-sample period. The LPL for real economic growth from period l_1 to l_2 is

$$LPL = \log \prod_{t=l_1}^{l_2} p(g_{t+1}|g_{1:t}, r_{1:t}) = \sum_{t=l_1}^{l_2} \log p(g_{t+1}|g_{1:t}, r_{1:t}) \tag{16}$$

The log-predictive Bayes factor is formed by subtracting the LPL of any two models. Values above or below 5 strongly favor one of the models.

A drawback is that Eq. (16) indicates overall forecast performance and conveys no information about performance during each out-of-sample period. For example, Model A often can surpass Model B in overall LPL, but B shows better accuracy of forecasts for a specified periods. We introduce the accumulated Bayes factor to address this concern. It is the following sequence of Bayes factors:

$$LPL_{l_1} = \sum_{t=l_1}^{l_2} \log p(g_{t+1}|g_{1:t}, r_{1:t}) \quad for \quad l_1 = 1, \ldots, T \tag{17}$$

This extension recursively shows Bayes factors for each out-of-sample period. By plotting their accumulation, we discern overall accuracy of forecasts[20], and recognize performance of time-varying forecasts. This is key for empirical time-series studies, as it reveals whether any model or lagged variable forecasts well and consistently over time.

5 Empirical Results

5.1 Data

We present data in a manner resembling that of Hamilton & Lin (1996) and Fama (1990), but our sample period is greatly extended. Each series embodies 1068 observations from February 1926 to December 2014. The monthly value-weighted return

[20]It happens to be the end point of the accumulated Bayes factor.

(includes dividend yield) of the S&P500 minus the three-month average of Fama's risk-free rate from the Center for Research in Security Prices are used to construct monthly excess market returns (r_t), our definition of "stock market returns". The industrial production index (IPI) is from the Federal Reserve Bank of St.Louis. Real economic growth (g_t) is the changes in the natural logarithm of the IPI. Stock market returns and real economic growth are scaled by 100. Table 3 illustrates their statistical summaries.

5.2 Model Priors

Since hierarchical priors are infeasible for AR, GARCH-AR, SV-AR and BEKK-VAR models[21], we select non-informative priors for them and let priors of $(c, a_1, \ldots, a_q, b_1, \ldots, b_p)$ be a multivariate normal distribution with means of 0 and variance-covariance of an identity matrix. The prior of σ in AR is a gamma distribution such as $\sigma_{t-1}^{-2} \sim Gamma(3, 1)$ with $E(\sigma^{-2} = 3)$.

The standard normal distribution is applied to the priors of ω, α and β in GARCH-AR[22]. Similarly, priors of ω, α follow the standard normal distribution of SV-AR. $\sigma_\nu \sim Gamma(3, 1)$ of SV-AR. For hyper-parameters among priors of VAR and BEKK-VAR, we chose non-informative priors such as $(\mathbf{c}, \mathbf{a_1}, \ldots, \mathbf{a_q}, \mathbf{b_1}, \ldots, \mathbf{b_p})$. They follow multivariate normal distributions with means of 0 and the variance-covariance of an identity matrix. Priors of Σ in VAR are Wishart distributions such as $\Sigma^{-1} \sim Wishart(2, I_2)$, where I_2 is a two-dimension identity matrix. Priors for each element in C, A, and B of BEKK-VAR follow an independent standard normal distribution with corresponding restrictions[23].

Hierarchical priors are a special feature of the Markov-switching model. Priors for ϑ are the same for both MS-AR, MS-VAR, IHMM-AR and IHMM-VAR except the dimension in a multivariate case is double that of an univariate case[24].

Let $\vartheta \sim MN(d_0, D_0)$, where d_0 is a vector of zeros, and D_0 is an identity matrix[25]. In like manner, $D_0^{-1} \sim Wishart(\tau_1, H_1)$, where $\tau_1 = 1+q+p$ for MS-AR and IHMM-AR models. $\tau_1 = 2 + q + p$ in MS2-VAR and IHMM-VAR models. H_1 is an identity matrix with corresponding dimension restrictions.

When models are MS2-AR and IHMM-AR, we let $\sigma^{-1} \sim Gamma(\nu_0, \lambda_0)$, where $\nu_0 \sim Gamma(3, 1)$ and $\lambda_0 \sim Gamma(3, 1)$[26]. While they are MS2-VAR and IHMM-

[21] Hierarchical priors require mixture models such as Markov-switching with two states. AR and VAR models implicitly assume one state.

[22] Priors for the autoregressive part of the equation are the same as AR. Restrictions for ω, α and β are that $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$.

[23] $C\prime$ is restricted to a lower diagonal matrix, and its diagonal elements are restricted to positive values. Diagonal elements of A and B are also restricted to positive to guarantee Σ is positive definite. A is a symmetric matrix and B a diagonal matrix.

[24] $\vartheta = (c, a_1, \ldots, a_q, b_1, \ldots, b_p)$ if models are MS-AR and IHMM-AR. $\vartheta = (\mathbf{c}, \mathbf{a_1}, \ldots, \mathbf{a_q}, \mathbf{b_1}, \ldots, \mathbf{b_p})$ if models are MS-VAR and IHMM-VAR.

[25] The dimension for MS-AR and IHMM-AR is $1 + q + p$. That for MS2-VAR and IHMM-VAR is $2 + 2q + 2p$.

[26] $E(\nu_0) = E(\lambda_0) = 3$.

VAR, we let $\Sigma \sim Wishart(\tau_0, H_1)$, where $\tau_0 \sim Gamma(3,1)I(\tau_0 \geq 2)$ and $H_0 \sim InvWishart(2, I)$. I is an identity matrix with two dimensions.

For hyper priors for η and α of IHMM-AR and IHMM-VAR, we let $\chi_1 = \chi_3 = 3$ and $\chi_2 = \chi_4 = 1$. Choice of priors, hierarchical priors, and hyper-prior is applied to both full sample estimation and out-of-sample forecasts.

5.3 Predictive Power of Stock Returns

Table 1 summarizes the predictive power of lagged stock returns on future real growth[27]. According to density forecasts, lagged stock returns significantly predict future real growth. We calculated the Bayes factor between models with and without lagged stock returns to reveal whether lagged stock returns predict economic growth.

To reach a robust outcome, we controlled lags on real growth within each model. Basically, entries underneath $b_{1:p} = 0$ (second column) exhibit the highest LPL among $q = 1, 2, 3, 4$ while $b_{1:p} = 0$ where $b_{1:p} = 0$ in each model. Entries beneath $b_{1:p} \neq 0$ (third column) indicate the highest LPL among lag combinations of $p = 1, 2, 3, 4$ and $q = 0, 1, 2, 3, 4$ in the interim that the lagged stock returns are introduced[28].

Entries for the Bayes factor indicate the significance of the predictive power of lagged stock returns, it is LPL when $b_{1:p} = 0$ minus LPL $b_{1:p} \neq 0$. Per Table 1, all benchmark models and our IHMM suggest that lagged stock returns predict economic growth. Our finding supports existing literature. Especially in top-performing models, the predictive power of lagged stock returns is considerately robust.

As Section 4, limitations arise from using Bayes factors and LPL. Figure 1 compares accumulated Bayes factors of models with and without lagged stock returns. The objective is to investigate whether the predictive power of lagged stock returns changes over time. Table 1 reveals several notable results and addresses concerns arising from Hamilton & Lin (1996)[29] and Choi et al. (1999)[30]. First, it documents dramatic gains in accuracy of forecasts driven by lagged stock returns from the late 1920s to the early 1940s. Substantial predictive power is present during the 1950s and post-2008. Second, predictive power erodes from the 1930s until the 1990s[31]. To confirm the time-varying feature of its predictive power, Table 2 shows the predictive power of lagged stock returns in assorted subsample. It affirms Figure 1 in that their predictive power changes with time. Strong and weak predictive power are documented in differing subsample periods. In regard to Hamilton & Lin (1996), our approach suggests IHMM is a flexible

[27]In our preliminary results, we investigate how lagged stock returns and economic growths affect predicting future real growth. For example, we compute the LPL of each benchmark model with four lags (four months) of stock returns and economic growth.

[28]Each entry represents the maximum LPL among 20 lag combinations of each model. For example, we calculate the LPL for AR for $p = 1, 2, 3, 4$ and $q = 0, 1, 2, 3, 4$. The maximum is -1893.1, which enters the table. Details of LPL results for each model (including RMSEF and LPL) are in the Appendix.

[29]It suggests that the predictive power of stock returns and changes in economic growths are driven by states.

[30]It suggests a weak predictive relationship.

[31]Finds are based on the top plot in the Figure 1.

mechanism that lets us model predictive power in a time-varying manner rather than representing economic conditions. Table 2 also explains why Choi et al. (1999) suggest a weak predictive relationship.

Our findings advance the literature in several ways. First, we examine the impacts of lagged stock returns through out-of-sample density forecasting, whereas previous works target only linear or nonlinear in-sample correlations or point forecasts[32]. Second, we investigate the predictive power of lagged stock returns under various advanced benchmark models and their lag structures, an aspect that earlier studies barely pay attention to. We discover that lag structures make a significant difference[33]. This makes sense in that the real economy often reacts more slowly than financial markets. Third, we propose a Bayesian nonparametric model IHMM to study their relationship. It soon is evident that our models surpass the forecast performance of existing benchmarks. What is most significant is our approach surpasses previous benchmarks in accommodating dynamics parameters and heteroscedasticity while lagged stock returns remain robust.

5.4 Model Selection

This section shows the robustness and accountability of employing IHMM-AR and IHMM-VAR to model the market-economy dynamic.

Although Table 1 affirms IHMM-AR and IHMM-VAR as the best model, it display only overall forecast performance across the entire out-of-sample period. To investigate whether benchmark models outperform them during some periods, we study accumulated Bayes factors.

Figure 2 displays accumulated Bayes factors for IHMM-AR and other benchmark including IHMM-VAR and a complete picture of forecasting performance. Versus benchmark models using AR, VAR and MS-AR, our model exhibits a consistent increasing trend. A similar trend is picked up by GARCH (top plot) with smaller magnitude than the previous plot. BEKK indicates a structural change during the mid-1940s. SV adn MS-AR illustrate minor gains pre-1970s, but are surpassed quickly thereafter. In sum, our approach consistently provides accurate forecasts.

As mentioned, the regime or states spaces of IHMM are fully determined by the data. When new data signal new states, IHMM adapts quickly with appropriate changes. For example, the rise in accumulated Bayes factors between IHMM-AR and MS-VAR accelerate after the early 1980s due to the Great Moderation. MS-VAR fails to accommodate such new feature due to its limited state spaces, whereas IHMM-AR can expand state spaces continually for any structural change[34]. Another example is Figure 3, poste-

[32]Fama (1990), Schwert (1990), Choi et al. (1999), Lee (1992), Hassapis & Kalyvitis (2002) and Kanas & Ioannidis (2010). The drawback of an in-sample relationship is that it cannot determine predictive power. Point forecasts cannot account for heteroscedasticity, as it measures the center of the predictive distribution.

[33]For details of results, see the Appendix.

[34]Similar patterns emerge by replacing IHMM-AR with IHMM-VAR. See the Appendix.

rior averages for the number of regimes by recursive samples. Between 1940 and 1970, the number of regimes decreases. It is a unique and attractive feature of IHMM thatregime spaces do not necessarily increase with sample size[35]. Similar patterns appear in IHMM-VAR with slower increasing paths than those in IHMM-AR.

Its flexible framework undoubtedly makes IHMM a competitive model, but an important question remains: where does its greater forecasting accuracy come from? Eq.(18) offers an answer: from the predictive density curve. Let the Θ denote the parameter space.

$$p(g_{t+1}|g_{1:t}, \Theta, s_t) = \sum_{k=1}^{\infty} \pi_{s_t k} N(g_{t+1}|a_{1:q,k}, b_{1:p,k}, \sigma_k, g_{1:t}) \tag{18}$$

Regardless of the form of target density $p(g_{t+1}|g_{1:t}, \Theta, s_t)$, IHMM always tries to construct the target density that best fits the data through the right side of the equation[36]. Figure 4 is a snapshot of the predictive likelihood computation. Each density curve represents the predictive density distribution on the data indicated at top.

Eq.(18) is displayed in Figure 4. The flexible frameworks of IHMM-AR and IHMM-VAR adjust the thickness of the tail and shift the center of the predictive distribution. For instance, IHMM for Oct. 1935 suggests thinner tails and a rightward shift of the center with respect to other benchmarks. Similar cases appear in the bottom left (Aug.1970), where IHMM suggests a leftward shift to model future likelihood rather than imposing a thick tail.

In summary, IHMM allows features from conventional Markov-switching models, where every parameter can be time-varying, and unconstrained by its state space. Its unrestricted state spaces let us explore volatile dynamics as well as GARCH, SV and BEKK models. Our model reveals both tail adjustments and the center shift in selected predictive distributions in Figure 4. Our IHMM approach delivers more accurate out-of-sample forecasts than benchmark models.

5.5 Parameter Dynamics of IHMM

This section investigates the benefits of IHMM through posterior analysis. As noted earlier, IHMM facilitates exploring target density by flexible approximation irrespective of its true form. This section illustrates how parameters behave through posterior analysis.

[35]The swift rise during the 1960s indicates the approaching Great Moderation.

[36]Conventional AR and VAR models do not reveal temporal dynamics; parameters $a_{1:q}, b_{1:p}, \sigma$ are always constant. GARCH, BEKK, and SV allow heteroscedasticity (σ_t), but $a_{1:q}, b_{1:p}$ are constant over time. Often, these heteroscedasticity models experience persist volatility, a drawback for modeling macroeconomic time series. Maheu & Yang (2016) show the poor performance of GARCH on US T-bill. MS-AR and MS-VAR with two states offer some flexibility, but it has been largely restricted to state space. Choosing an appropriate state space and keeping state spaces dynamic have been challenging for conventional MS models. The IHMM family offers a solution for these issues.

Figure 5 and 6 show posterior parameter averages for IHMM-VAR, VAR, BEKK and MS-VAR[37]. IHMM-VAR captures the same structural changes for b_1, b_2, and b_3 during the 1960s as IHMM-AR but other benchmarks show little change. Given IHMM describes the data better than other benchmarks, it captures parameter dynamics that more likely are closet to the true data generating process. IHMM is distinguished from fixed-state Markov-switching models in two ways: its state spaces are unbounded, and it is expandable when necessary. State spaces in MS-AR and MS-VAR models are predetermined before estimation. As a result, IHMM implicitly samples state spaces as posterior distributions like any other parameter.

Figure 9 illustrates posterior distributions of the number of states. IHMM-AR averages 5 to 6 states and IHMM-VAR 8-9. States number is sensitive to choices of hyper parameters in Eq. (9), but out-of-sample performance is not[38]. There are always cses in which 1 or 2 observations are assigned to a newly introduced state (often they are data outliers), whereas a loose hyper-prior in Eq. (9) frequently introduces new states to the pool. Due to their trivial assignment, these states are uncertain[39], because their observations can be absorbed suddenly into large active states. However, these small states are weighted equally with large active states, creating the illusion the data demand a multitude of states. They actually have no impact when we take into account the density forecast due to its trivial scale.

Table 5 reveals results of sensitivity tests on different choices of parameters in Eq. (9). LPL hardly changes by imposing different degrees of tightness on hyper-priors.

Figure 10 is a heat map of IHMM-AR[40]. Heat maps are another way to express the structural dynamics of IHMM. A heat map is a square matrix with a sample size indicated on the x and y-axis as its dimension. They are dated February 1926 through December 2014. Each cell represents the corresponding two dates. Over each MCMC iteration, we filled out the matrix by discerning if two dates share the same state. When sharing happens, the corresponding cell will have an increment. In the next MCMC iteration, we repeated the above steps and summed the counts in each cell. In the end, cells bear a number corresponds to the number of MCMC iteration (yellow on map), suggesting corresponding two dates always share the same states through all iterations. Other cells may read zero (red on map), which suggests the corresponding two dates share hardly anything in common. Then, we divided each cell by the total of MCMC iterations to arrive at Figure 10, which shows probability that any two dates share the same state[41].

We observe a structural change in the early 1960s because later years share hardly

[37]We choose$q = 3$ and $p = 3$ for IHMM-AR, AR, GARCH, SV, and MS-AR. Their LPLs are highest among all lag specifications.

[38]In the robustness check, we show the sensitivity test regarding choices of hyper-prior parameters.

[39]Van Gael et al. (2008) use the ploy urn scheme to explain it.

[40]We show the heat map generated by IHMM-AR of $q = 3$ and $p = 3$. The heat map for IHMM-VAR closely resembles IHMM-AR, so, we do not discuss it. It is available in the Appendix

[41]The heat map is symmetrical around the diagonal. Yellow indicates the probability of closing to 1.

any state that occurred pre-1960. Kim & Nelson (1999) document the Great Moderation in 1984 using MS-AR, but our model declares it actually occurred during the early 1960s. The parameter changes appear considerably more dynamic from 1960s to early 1980 than after 1984, but the heat map does not distinguish them from each. This finding would reward further research. In sum, IHMM accommodates heteroscedasticity and time-vaying parameters in a unified framework.

6 Robustness Study

6.1 Sticky Prior Extension

Fox et al. (2011) introduced the sticky version of IHMM. It is a prior that favors more weights on self-transition states. The Eq. (8) is rewritten as follows:

$$\Pi_j|\alpha,\Gamma \overset{iid}{\sim} Stick2\Big(\alpha+\kappa, \frac{\alpha\Gamma+\kappa\delta_j}{\alpha+\kappa}\Big), \qquad j=1,2,\ldots,$$

This indicates that element j of row Π_j has a base distribution of $\frac{\alpha\Gamma+\kappa\delta_j}{\alpha+\kappa}$ whereas the base distribution of other elements in jth row is $\frac{\alpha\Gamma}{\alpha+\kappa}$. As a result, when $\kappa > 0$, strong preference is given to the self-transition probability since $E[\pi_{jj}] = \frac{\alpha\Gamma+\kappa\delta_j}{\alpha+\kappa}$. The sticky version easily reverts to the non-sticky version by letting $\kappa = 0$. To study how sticky priors affect model performance, we assign different values of κ[42] and test different selections of lagged returns. Per Table 4, the LPL of IHMM under different choices of κ do not alter our conclusion. One reason could be that the large sample period is exceptionally dynamic, and imposing a prior on self-transition is not a rewarding way to exploring them.

6.2 Sensitivity Test of Dirichlet Process

An inappropriate choice of priors often affect results of Bayesian models. As noted, α and η are key elements that govern how frequently new parameters are introduced. To sustain data-driven results, we set hyper-priors on α and η as indicated by Eq. (9).

Table 5 summarizes LPLs of IHMM-AR and IHMM-VAR under different selections of hyper parameters. Per table, combinations of hyper-prior do not contribute significantly to LPL. One reason is the addition of the hyper-prior. Another may be that the LPL computation leaves the first 20 observations as a training sample, which likely makes LPL indifferent to by the prior when sample size is small[43].

6.3 Term and Credit Spreads

Faust et al. (2013) show that their chosen credit spreads indicate substantial predictive

[42] $\kappa = 0.8, 1.5, 3.0$

[43] This also is the main reason we do not compute marginal likelihood.

power for future growth. We explore the predictive power of lagged monthly credit and term spreads on real growth[44].

Table 6 summarizes term and credit spreads. We apply IHMM to the new dataset and compare with stock returns. Again, we investigate the predictive power of various lag of term and credit spreads on real growth.

Table 7 summarizes forecasts under each dataset. It indicates no significant differences in forecasting real growth by adding lags or credit spreads. This finding contravenes Faust et al. (2013), who insist credit spreads significantly predict economic growth. There are several possible reasons. We use credit and term spreads to predict montly industrial production as our measure of economic growth; they use quarterly real GDP. Also, they use Bayesian Model Averaging and AR to study the predictive power of spreads, whereas we invoke IHMM. Lagged credit spreads could explain many dynamics in simple models, but those dynamics likely are captured by states rather than other lagged variables using IHMM. Divergences between our work and that Faust et al. (2013) suggest directions for further study.

7 Conclusion

This paper introduced a Bayesian nonparametric model to study relations between stock market returns and real economic growth. That new approach significantly improved the accuracy of out-of-sample density forecasts through the lens of lag structures studies and benchmark models comparisons. Extensive results indicated our approach provides robust evidence that lagged stock returns significantly predict economic growth in a time-varying manner.

[44]Term spreads are from the Federal Reserve Bank of St.Louis. Spreads are yields on 10-year T-bill yield minus yields on three-month T-Bills. Credit spreads are from Prof. Simon Gilchrist's website http://people.bu.edu/sgilchri/. The sample period of spreads spans from 1982 to 2014.

References

Albert, J. H. & Chib, S. (1993), 'Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts', *Journal of Business & Economic Statistics* **11**(1), 1–15.

Beal, M., Ghahramani, Z. & Rasmussen, C. (2002), 'The infinite hidden Markov model', *In Advances in Neural Information Processing Systems* pp. 577–584.

Choi, J., Hauser, S. & Kopecky, K. (1999), 'Does the stock market predict real activity? Time series evidence from the G-7 countries', *Journal of banking and finance* **23**, 1771–1792.

Dufays, A. (2015), 'Infinite-state Markov-switching for dynamic volatility', *Journal of Financial Econometrics* .

Fama, E. (1990), 'Stock returns, expected returns, and real activity', *Journal of Finance* **4**, 1089–1108.

Faust, J.and Gilchrist, S., Wright, J. & Zakrajsek, E. (2013), 'Credit spreads as predictors of real-time economic activity: a Bayesian model-average approach', *The Review of Economics and Statistics* **5**, 1501–1519.

Fox, E., Sudderth, E., Jordan, M. & Willsky, A. (2011), 'A sticky HDP-HMM with application to speaker diarization', *Annals of Applied Statistics* **5**, 1020–1056.

Hamilton, J. & Lin, G. (1996), 'Stock market volatility and the business cycle', *Journal of Applied Econometrics* **1**, 573–593.

Hassapis, C. & Kalyvitis, S. (2002), 'Investigating the links between growth and real stock price changes with empirical evidence from the G-7 economies', *The Quarterly Review of Economics and Finance* **42**, 543–575.

Hou, C. (2017), 'Infinite hidden Markov-switching vars with application to macroeconomic forecast', *International Journal of Forecasting* **33**, 1025–1043.

Jensen, M. & Maheu, M. (2010), 'Bayesian semiparametric stochastic volatility modeling', *Journal of Econometrics* **157**, 306–316.

Jensen, M. & Maheu, M. (2013), 'Bayesian semiparametric multivariate GARCH modeling', *Journal of Econometrics* **176**, 3–17.

Jochmann, M. (2015), 'Modeling U.S. inflation dynamics: a Bayesian nonparametric approach', *Econometric Reviews* **34**.

Kanas, A. & Ioannidis, C. (2010), 'Causality from real stock returns to real activity: evidence of regime-dependence', *International Journal of Finance and Economics* **15**, 180–197.

Kim, C. & Nelson, C. (1999), 'Has the US economy become more stable? a Bayesian approach based on a Markov-switching model of the business cycle', *The Review of Economics and Statistics* **81**, 608–616.

Kim, S. & In, F. (2003), 'The relationship between financial variables and real economic activity: evidence from spectral and wavelet analyses', *Studies in Nonlinear Dynamics and Econometrics* **7**.

Koop, G. & Korobilis, D. (2009), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics* **3**, 267–358.

Lee, B. (1992), 'Causal relations among stock returns, interest rates, real activity, and inflation', *The Journal of Finance* **48**, 1591–1604.

Maheu, M. & Yang, Q. (2016), 'An infinite hidden Markov model for short-term interest rates', *Journal of Empirical Finance* **38**, 202–220.

Schwert, G. (1990), 'Stock returns and real economic activity: A century of evidence', *Journal of Finance* **45**, 1237–1257.

Sethuraman, J. (1994), 'A constructive definition of Dirichlet priors', *Statistica Sinica* **4**, 639–650.

Shi, S. & Song, Y. (2015), 'Identifying speculative bubbles using an infinite hidden Markov model', *Journal of Financial Econometrics* .

Song, Y. (2013), 'Modelling regime switching and structural breaks with an infinite hidden Markov model', *Journal of Applied Econometrics* **29**.

Teh, Y., Jordan, M., Beal, M. & Blei, D. (2006), 'Hierarchical Dirichlet processes', *Journal of the American Statistical Association* **101**, 1566–1581.

Van Gael, J., Saatci, Y., Teh, Y. & Ghahramani, Z. (2008), Beam sampling for the infinite hidden Markov model, *in* 'Proceedings of the 25th International Conference on Machine Learning:', pp. 1088–1095.

Table 1: Summary of Predictive Power of Lagged Stock Returns

Model	$b_{1:p} = 0$	$b_{1:p} \neq 0$	Bayes Factor
AR	-1941.2	-1893.1	48.1
MS-AR	-1932.4	-1864.2	68.2
GARCH-AR	-1524.9	-1512.3	12.6
SV-AR	-1488.8	-1477.1	11.7
VAR	-1936.2	-1887.1	49.1
MS-VAR	-1523.8	-1510.1	13.7
BEKK-VAR	-1483.0	-1472.6	10.4
IHMM-AR	-1458.8	-1446.5	12.3
IHMM-VAR	-1457.2	-1451.1	6.1

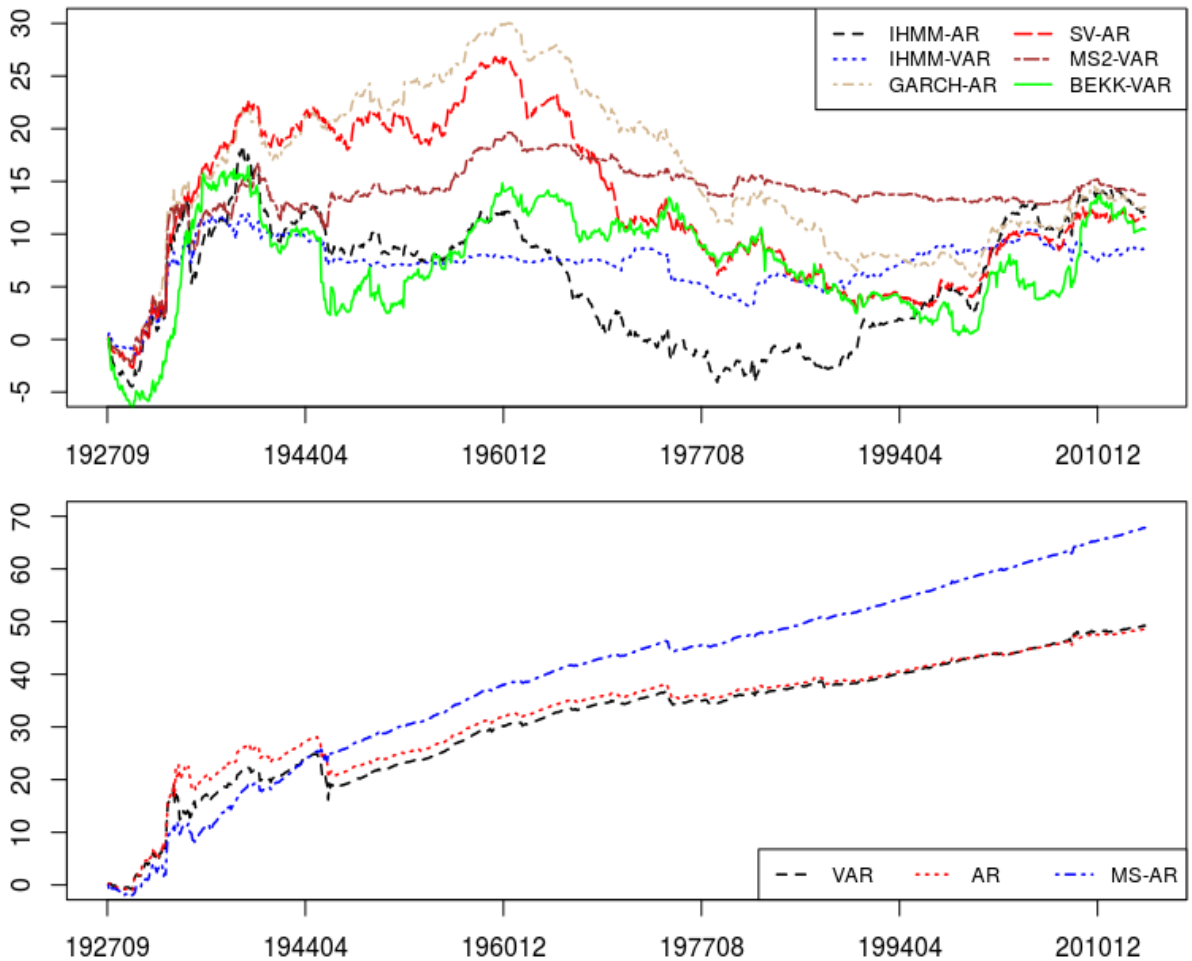
Sample size is 1068 and out-of-sample size is 1048. The Bayes factor is calculated by subtracting LPL($b_{1:p} = 0$) from LPL($b_{1:p} \neq 0$). Bayes factors exceeding 5, implies lagged stock returns have predictive power for future real growth. The second column indicates the highest LPL of each model when stock returns are excluded. The third column suggests the highest LPL when stock returns are included. The Appendix gives details of the LPL performance of different lag structures for each model.

Table 2: Bayes Factors of Indicating Predictive Power of Lagged Stock Returns in Subsamples

Model	$192709 - 194001$	$194002 - 195101$	$195102 - 198001$	$198002 - 201512$
AR	26.04	-2.11	12.37	12.25
MS-AR	19.25	9.55	17.60	21.40
GARCH-AR	21.04	0.71	-10.48	1.27
SV-AR	22.09	-1.92	-11.93	3.37
VAR	21.89	0.08	13.36	13.90
MS-VAR	15.92	-2.05	-0.15	-0.02
BEKK-VAR	13.83	-11.10	5.37	2.35
IHMM-AR	16.09	-7.84	-11.22	15.04
IHMM-VAR	10.22	-4.04	-2.96	4.13

Each entry represents the Bayes factor of LPL($b_{1:p} = 0$) minus LPL($b_{1:p} \neq 0$) for each corresponding model in the first column. The out-of-sample period is in the first row. Bayes factor is exceeding 5 imply lagged stock returns have predictive power for future real growth in the corresponding subsample period. The Appendix provides details of LPL performance of different lag structures for each model.

Figure 1: Accumulated Bayes Factors between Models with and without Lagged Returns



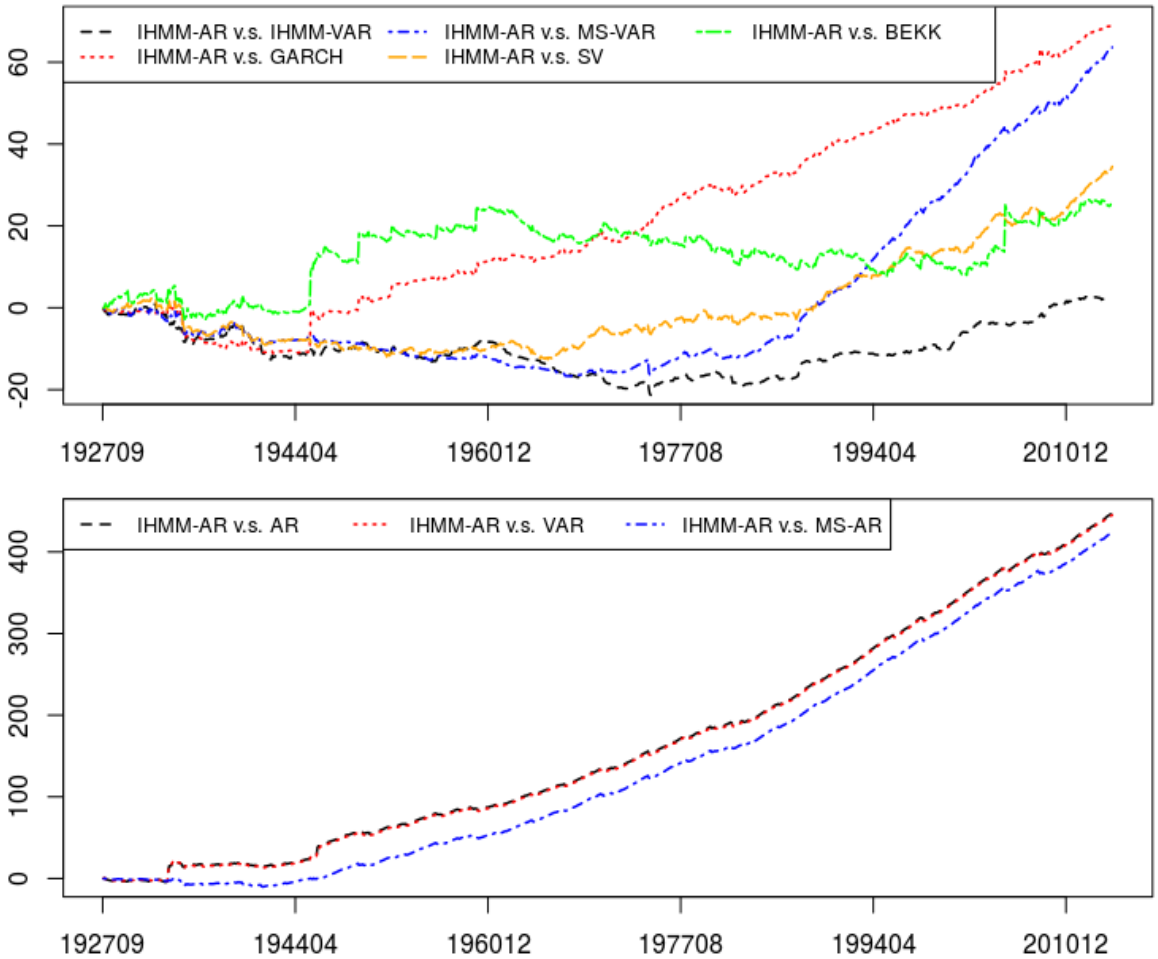
The top (bottom) plot recounts accumulated Bayes factors for values below (above) than 30. All models are set at $q = 3$. All are lagged stock returns are set at $p = 3$.

Table 3: Statistical Summaries of Stock Returns and Real Economic Growth

Name	Mean	Variance	Median	25%Q	75%Q	Skewness	Kurtosis
Stock Returns	0.648	30.24	0.945	-2.043	3.595	0.409	9.603
Real Growth	0.263	3.211	0.308	-0.304	0.846	0.354	14.87

This table reports summary statistics for monthly excess returns on the S&P 500 stock excess returns and U.S. industrial production growth rates from February 1926 to December 2014 (1068 observations).

Figure 2: Accumulated Bayes Factor between IHMM-AR and Benchmarks



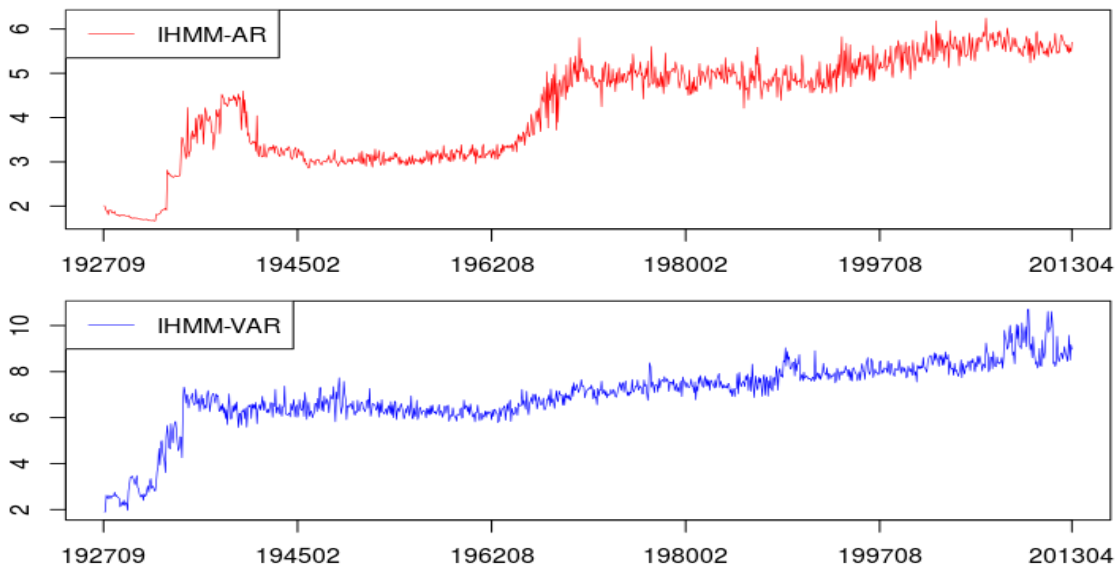
The top plot is the accumulated Bayes factor between IHMM-AR and models, which Bayes factors are below than 60. The bottom plot is Bayes factors above 60. Lags are restricted to $q = 3$ and $p = 3$ for all univariate models. Lag are set as $q = 3$ and $p = 1$ for multivariate case.

Table 4: Robustness Check on Sticky-Prior

	IHMM-AR			IHMM-VAR		
	$\kappa = 0.8$	$\kappa = 1.5$	$\kappa = 3.0$	$\kappa = 0.8$	$\kappa = 1.5$	$\kappa = 3.0$
$p = 0$	-1457.2	-1475.7	-1463.4	-1453.0	-1457.2	-1462.4
$p = 1$	-1453.4	-1462.0	-1454.7	-1451.9	1456.7	-1456.2
$p = 2$	-1453.9	-1454.4	-1456.1	-1458.1	-1458.2	-1458.5
$p = 3$	-1447.8	-1451.0	-1447.9	-1450.6	-1454.8	-1453.5
$p = 4$	-1458.8	-1460.1	-1460.5	-1464.9	1466.6	-1468.1

Sample size is 1068. Out-of-sample size is 1048. Entries in the table indicates LPL at different sticky-prior κ and lags of stock returns of $q = 3$ is imposed.

Figure 3: Posterior Average of Regimes under Recursive Samples



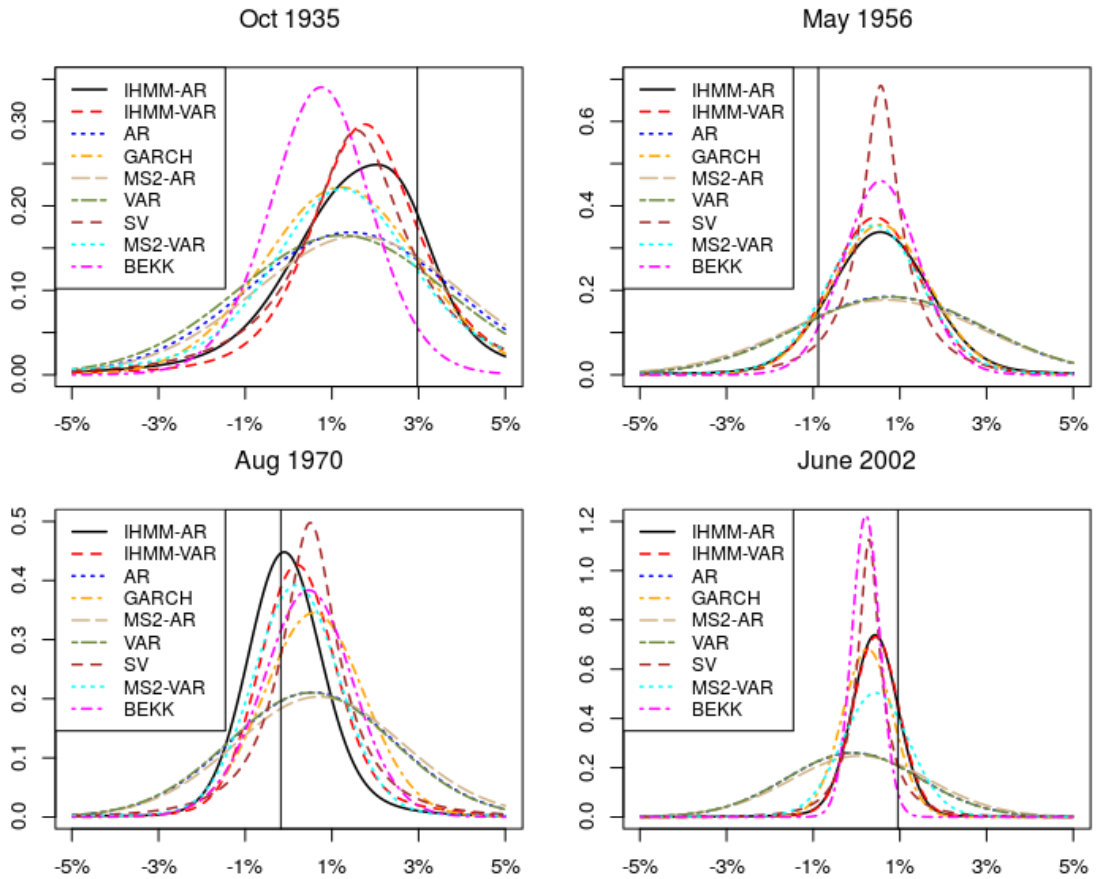
The plots are calculated through recursive expanding samples. Each value on the plot represents the posterior average of regime number using the information upon to the time corresponds to the date on the x-axis.

Table 5: Prior Sensitivity Test on Dirichlet Process

	IHMM-AR	
	$\alpha \sim Gamma(a_1, b_1)$ $\quad \eta \sim Gamma(a_2, b_2)$	
Very Loose	$a_1 = 5, b_1 = 1$ and $a_2 = 5$, $b_2 = 1$	-1448.3
Less Loose	$a_1 = 5, b_1 = 1$ and $a_2 = 2.5$, $b_2 = 0.5$	-1448.4
Regular	$a_1 = 2, b_1 = 1$ and $a_2 = 5$, $b_2 = 1$	-1444.6
Less Tight	$a_1 = 5, b_1 = 1$ and $a_2 = 2$, $b_2 = 8$	-1453.8
Very Tight	$a_1 = 2, b_1 = 2$ and $a_2 = 2$, $b_2 = 8$	-1449.2
	IHMM-VAR	
	$\alpha \sim Gamma(a_1, b_1)$ $\quad \eta \sim Gamma(a_2, b_2)$	
Very Loose	$a_1 = 5, b_1 = 1$ and $a_2 = 5$, $b_2 = 1$	-1453.6
Less Loose	$a_1 = 5, b_1 = 1$ and $a_2 = 2.5$, $b_2 = 0.5$	-1454.8
Regular	$a_1 = 2, b_1 = 1$ and $a_2 = 5$, $b_2 = 1$	-1448.4
Less Tight	$a_1 = 5, b_1 = 1$ and $a_2 = 2$, $b_2 = 8$	-1452.3
Very Tight	$a_1 = 2, b_1 = 2$ and $a_2 = 2$, $b_2 = 8$	-1450.8

Sample size is 1068. Out-of-sample size is 1048. Entries of last column indicates the LPL at different choices of hyper-priors on α and η. Lags are restricted to $q = 3$, $p = 3$ for IHMM-AR and $q = 3$, $p = 1$ for IHMM-VAR.

Figure 4: Predictive Density Curve on Selected Dates



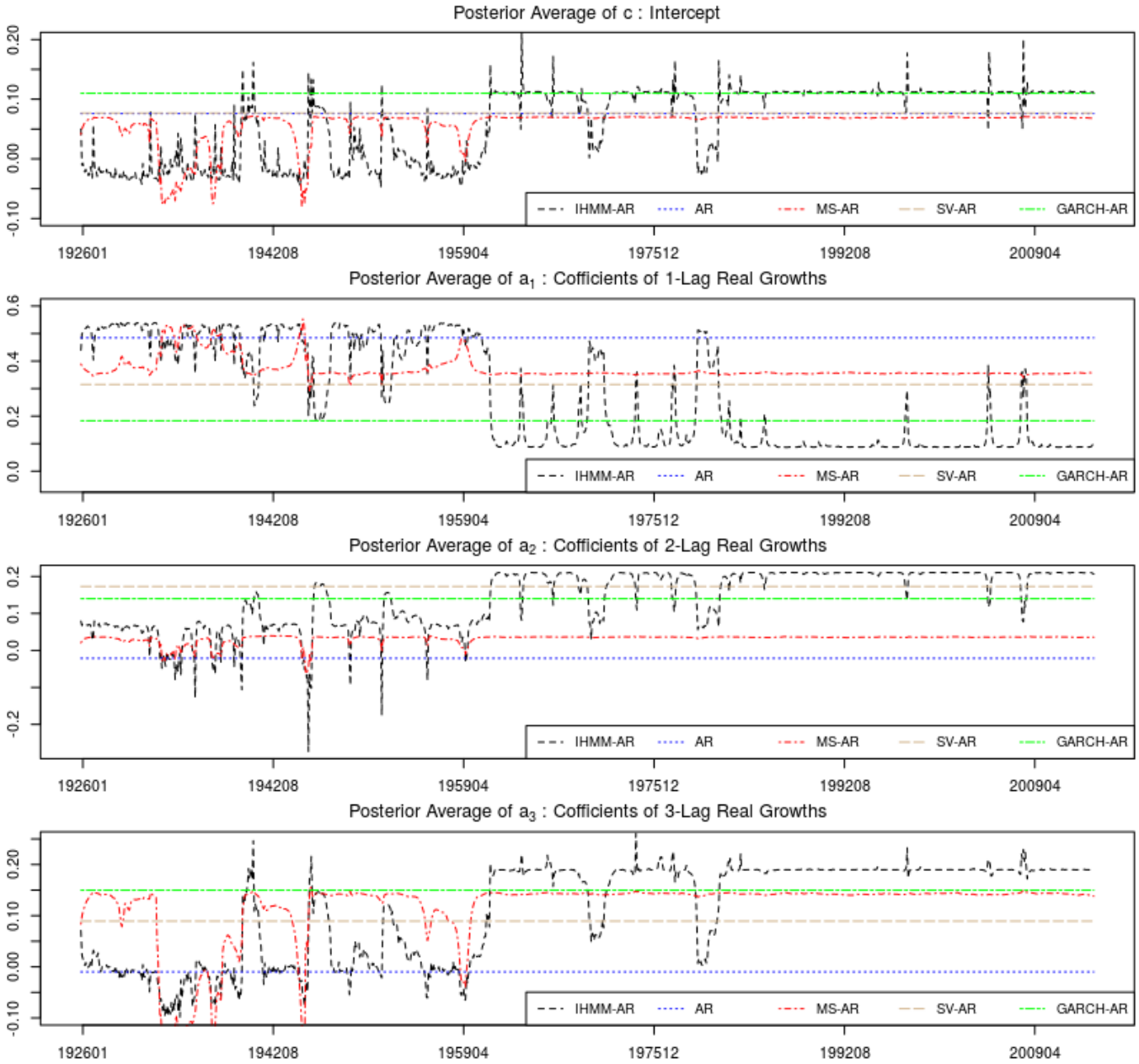
Lag specifications are restricted to $q = 3$ and $p = 3$ for all univariate models. Lag are set as $q = 3$ and $p = 1$ for multivariate cases.

Table 6: Statistical Summaries of Term and Credit Spreads

Name	Mean	Variance	Median	25%Q	75%Q	Skewness	Kurtosis
Term Spreads	1.064	0.212	0.940	0.760	1.240	2.108	8.721
Credits Spreads	1.865	1.284	2.040	0.920	2.772	-0.298	2.039

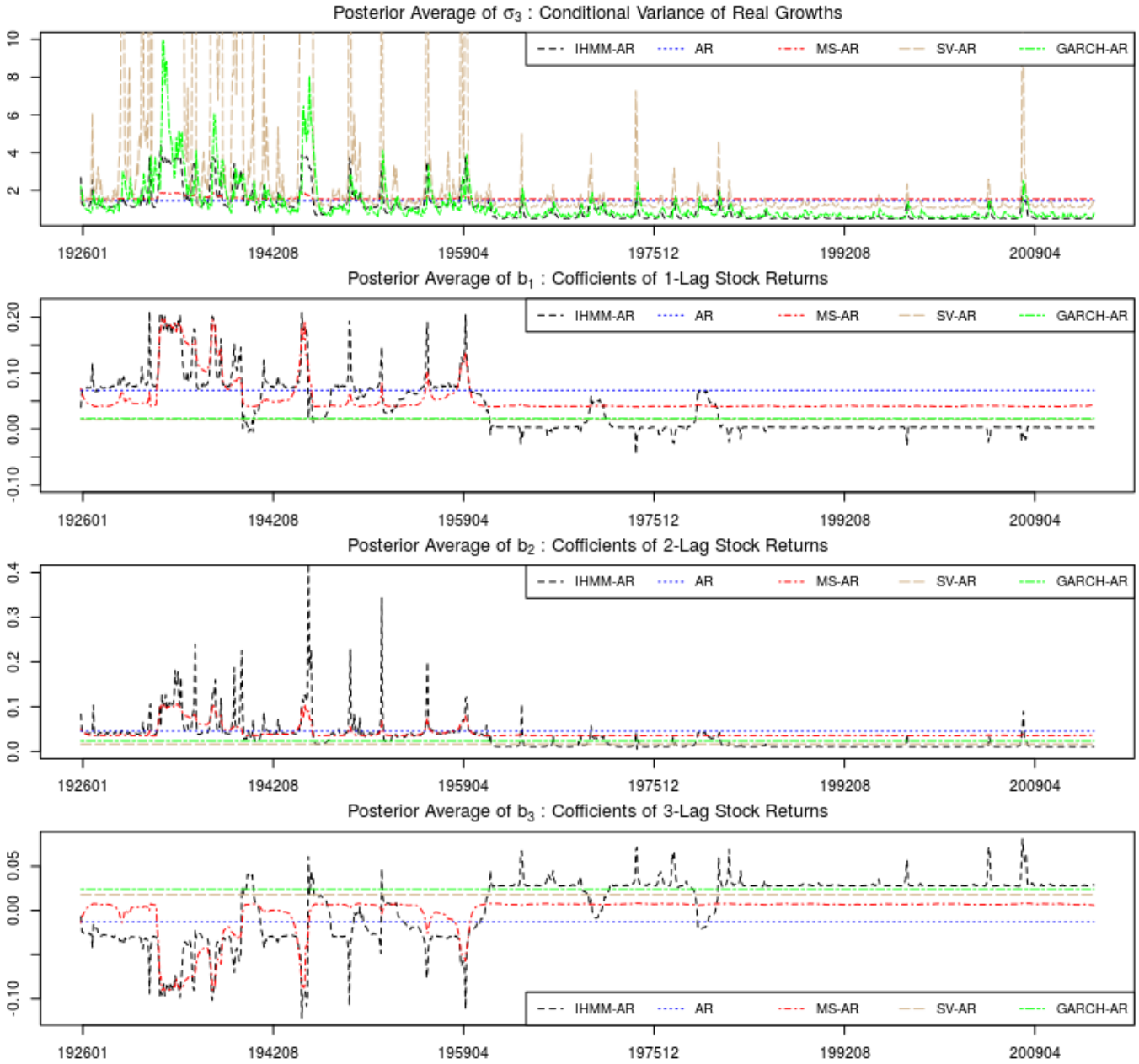
It is based on monthly term and credit spreads in percent spinning January 1982 to December 2014 (396 observations).

Figure 5: Parameter Posterior Average of Univariate Models (Part-One)



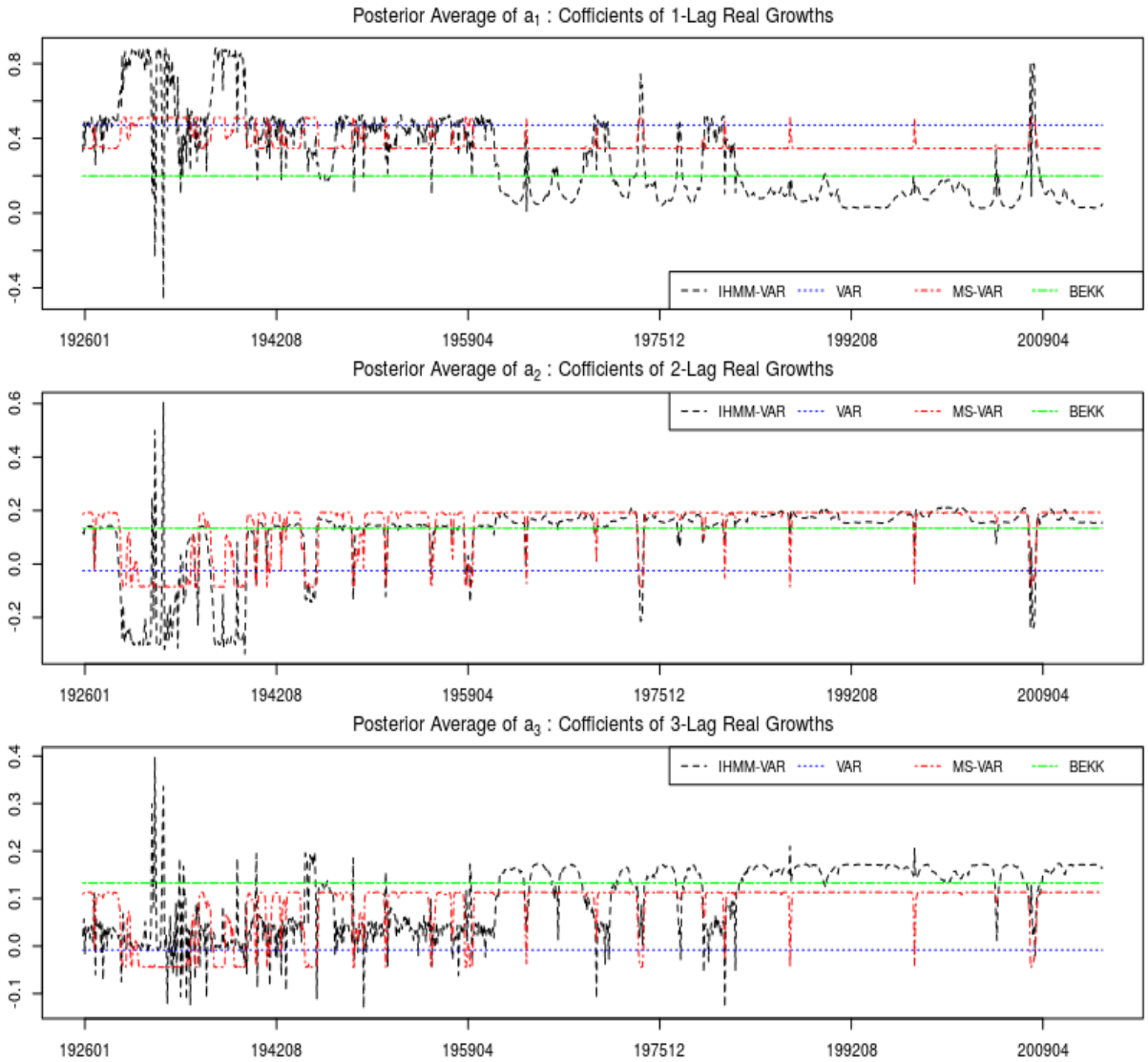
Posterior averages are drawn from the full sample of 1068. Corresponding lags are $q = 3$ and $p = 3$. Section 2 provides model details.

Figure 6: Parameter Posterior Average of Univariate Models (Part-Two)



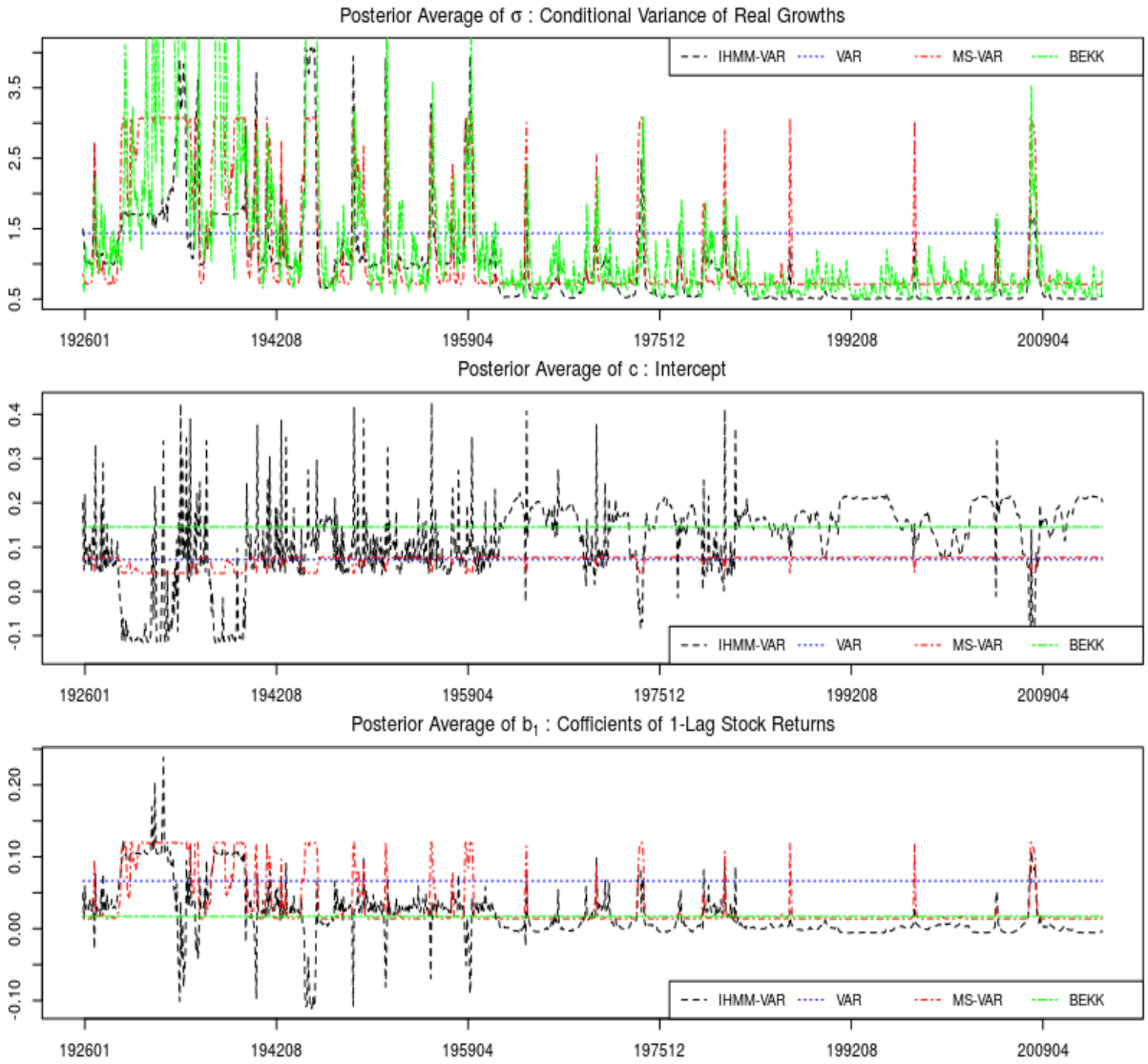
Posterior averages are drawn from the full sample of 1068. Corresponding lags are $q = 3$ and $p = 3$. Section 2 provides model details.

Figure 7: Parameter Posterior Average of Multivariate Models (Part-One)



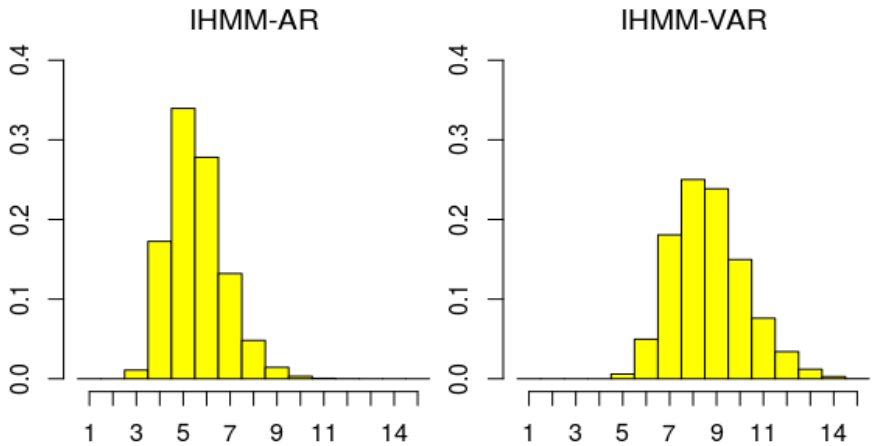
Posterior averages are drawn from the full sample of 1068. Corresponding lags are $q = 3$ and $p = 1$. Section 2 provides model details.

Figure 8: Parameter Posterior Average of Multivariate Models (Part-Two)



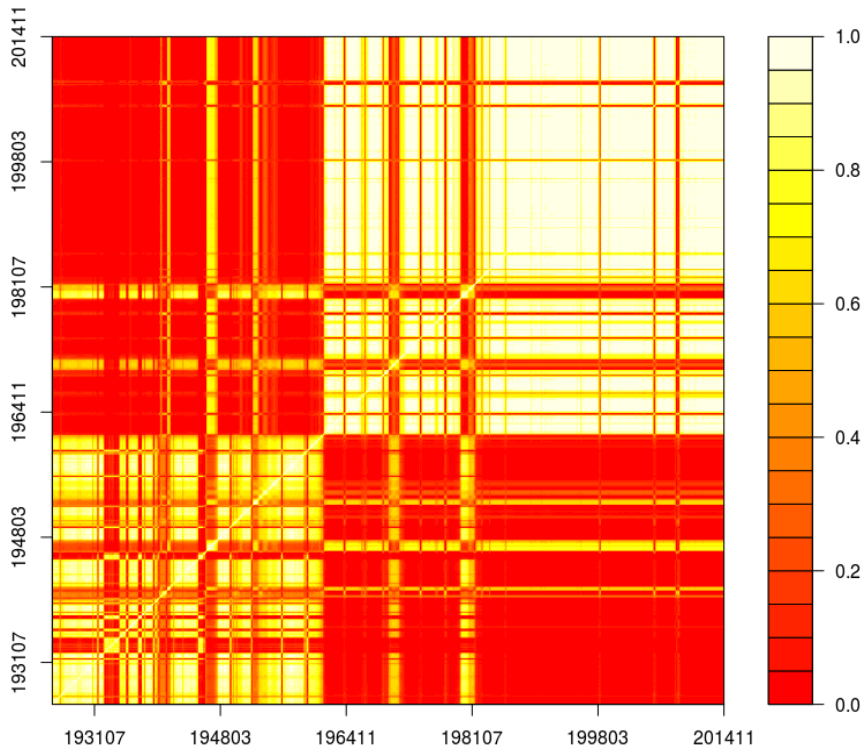
The posterior averages are sampled based on a full sample of 1068. The corresponding lag structure is $q = 3$ and $p = 1$. Model details are referred to Section 2.

Figure 9: Posterior Histogram of the Number of States



The posterior of states number is sampled at q=3&p=3 of IHMM-AR and q=3&p=1 of IHMM-VAR. The sample size 1068. The hyper-prior is $\alpha \sim Gamma(2,1)$ and $\eta \sim Gamma(2,1)$.

Figure 10: Heat Map of IHMM-AR



The heat map is sampled according to full sample. It shows the likelihood of any two dates share the same states.

Table 7: Predictive Power Investigation of Returns, Credits, and Bond

	Returns		Terms		Credits	
	IHMM-AR	IHMM-VAR	IHMM-AR	IHMM-VAR	IHMM-AR	IHMM-VAR
$p = 0$	-296.9	-304.3	-296.9	-310.2	-296.9	-319.5
$p = 1$	-304.0	-311.9	-300.0	-314.4	-300.5	-310.5
$p = 2$	-304.8	-313.7	-299.2	-313.7	-301.8	-312.2
$p = 3$	-299.9	-307.6	-294.9	-308.1	-305.4	-315.3
$p = 4$	-299.8	-308.7	-297.8	-310.6	-306.6	-316.3

Sample size is 396. Out-of-sample size is 376. The sample period for credit spreads is 1982. Each entry indicates the LPL at the corresponding specification when $q = 3$. Three datasets (first row) are used to predict future real growth. Corresponding models are indicated in the second row. The first column enumerate number of lags for returns, bonds, and credit.